Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
DTB Project: A Behavioral Model for Detecting Insider Threats	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **George Mason University,4400 University Drive,Fairfax,VA,22030-4444**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**2**	

DTB Project: A Behavioral Model for Detecting Insider Threats

Paulo C. G. Costa, Daniel Barbará, Kathryn B. Laskey, Edward J. Wright*, Ghazi Alghamdi, Sepideh Mirza, Mehul Revankar, Thomas Shackelford

George Mason University
4400 University Drive
Fairfax, VA, 22030-4444, USA
[pcosta, dbarbara, klaskey, galghamd, smirza, mrevanka, tschakel]
@gmu.edu

* Information Extraction and Transport, Inc.
1911 North Fort Myer, Suite 600
Arlington, VA, 22209, USA
ewright@iet.com

Keywords: Methods for Counter Denial and Deception, Novel Analysis Methods, Multi-Entity Bayesian Networks, Data Mining, Document Relevance, Behavior Modeling and Simulation, All Source, Counter Intelligence

Abstract

This paper describes the Detection of Threat Behavior (DTB) project, a joint effort being conducted by George Mason University (GMU) and Information Extraction and Transport, Inc. (IET). DTB uses novel approaches for detecting insiders in tightly controlled computing environments. Innovations include a distributed system of dynamically generated document-centric intelligent agents for document control, object-oriented hybrid logic-based and probabilistic modeling to characterize and detect illicit insider behaviors, and automated data collection and data mining of the operational environment to continually learn and update the underlying statistical and probabilistic nature of characteristic behaviors. To evaluate the DTB concept, we are conducting a human subjects experiment, which we will also include in our discussion.

1. Introduction

The overall idea of the DTB project is to model user queries and detect situations in which users in sensitive positions may be accessing documents outside their assigned areas of responsibility. This novel approach to detecting insider threats assumes a controlled environment in which rules for accessing information are clearly defined and, ideally, tightly enforced.

Although such environments provide little encouragement to insider threats, unusual access patterns are not easily perceived. In fact, documented cases in which insiders using unsophisticated tactics to outsmart standard security systems (CNN.com 1998, 2001) leave a very uncomfortable open question: how about the sophisticated ones?

Catching more elaborate patterns that might be characteristic of users attempting illegal activities such as disclosure of classified information is a daunting task that we tackle with a powerful inference method. The flexible modeling framework provided by multi-entity Bayesian networks (MEBN) makes it a natural candidate for modeling this complex problem.

2. Multi-Entity Bayesian Networks

MEBN logic (Laskey 2004) integrates First Order Predicate Calculus with Bayesian probability. It expresses probabilistic knowledge as a collection of MEBN fragments (MFrags) organized into MEBN Theories (MTheories). An MFrag represents a conditional probability distribution of the instances of its resident random variables given the values of instances of their parents in the Fragment graphs and given the context constraints.

Currently, MEBN logic is being implemented by IET's Quiddity*Suite™, a knowledge-based probabilistic reasoning toolkit, which we used to implement our models (see Alghamdi et al. 2004 for an initial discussion on the modeling efforts).

3. DTB Architecture

Initially, information on queries and overall system usage (e.g. document accesses, login times, copy and paste operations, etc.) is collected and stored into a generic Data Store, whereas specific data regarding the user's queries goes to the Wolfie data store.

Wolfie is a Data Mining application that assesses the relevance of each user query to his/her assigned task. Finally, the results from Wolfie are used as evidence to feed the Insider Bayesian Network model (IBN), a collection of MFrags (i.e. an MTheory) written in Quiddity*Suite™ and stored in a MFrags Knowledge

Funding for the DTB Project is provided by the Advanced Research and Development Activity (ARDA), under contract NBCHC030059, issued by the Department of the Interior. Additional support has also being provided by the US Navy.

Base. IBN is a MEBN behavioral model that uses the evidence provided on each user to assess the likelihood that his/her behavior patterns over a series of sessions is an indicative of malicious intent.

In order to integrate those different applications, we developed a Data Integration Module (DINT), which controls the information flow within the DTB architecture.

4. Interoperability

Our concept is intended to deal with a community with many possible users, both inside the Intel community and outside it. Like most complex domains, the Intelligence community does not have a commonly accepted conceptualization of its rules, policies, or vocabulary, making any attempt to build an interoperable model very difficult. As a simple example, how would our model cope with cases in which different organizations have different names for the same concept? Also, agencies may have different security and access policies. As an example, in some agencies access to USB ports and floppy disks is permitted, while in others the use of such devices is a sure passport to indictment.

Our approach to these and similar issues is the use of Ontologies, a modeling technique that formalizes the semantics of the domain being modeled. By providing formal representations of semantics, ontologies provide a uniform way to communicate our vision and to adapt our technology to new organizations and concepts. We developed two ontologies in parallel. The first, the *Insider behavior ontology* (*IB*), describes the MEBN model of an insider threat behavior; while the second ontology, the *Organization and Task Ontology* (*OT*), portrays the various aspects of an internal organization. For future implementations, we would just update these two ontologies to match the specific characteristics of the user agency.

5. Model Evaluation Strategy

A novel approach demands carefully designed evaluation. In our case, we started by exposing our behavioral model to domain specialists from outside our team, who suggested new aspects to be addressed and helped in fine-tuning our prior probabilities. After achieving a model that satisfied our external evaluators, we proceeded with simulation and sensitivity analysis experiments, in which we varied some of the parameters in order to analyze the overall robustness of our model.

Finally, a system designed to monitor human users should be evaluated with actual human users. For this purpose, we currently are conducting an experimental evaluation involving students from George Mason University's School of Information Technology and Engineering. Subjects perform research and analysis in an environment designed to mimic the target environment for our system. In addition to their overt task, some of the subjects are given a "clandestine" task involving a topic different from their assigned task. The purpose of the experiment is to evaluate how well our system can detect the subjects who are performing a clandestine task.

The results of both the computer and human subject experiments are expressed in terms of *probability of detection (PD)* and *probability of false alarms (PFA)*. The probability of detection is the probability of correctly detecting a threat behavior, while a false alarm happens when we declare a user to be a *Threat* while he or she is *Normal*.

By varying the threshold, PD and PFA can be traded off against each other. A useful tool for evaluating classifiers is the receiver operating characteristic (ROC) curve, which plots PD against PFA. The area under the ROC curve (AUC) is a threshold-independent measure of the quality of a classifier. The ROC curves along with the AUC's are used as the basis for analyzing the results of the computational simulation, sensitivity analysis, and the live subjects' experiments.

6. Conclusion

Although standard access control methods provide some measure of control against insider abuse, more protection is required. The potentially disastrous consequences of even a single successful breach argue for the development of more sophisticated methods of detecting malicious insiders. Strictly enforced policies are highly efficient to prevent such criminals to operate, but experience proves that relying on this alone is just not enough.

We presented a novel approach to insider threat detection, which we believe has the potential to greatly increase the efficacy and efficiency of current systems. Uncovering improper human behavior along a series of events was an intractable approach that is now feasible due to the recent advances in the field of Bayesian inference technology. The main contribution of the DTB project is to transform those advances into reliable applications for the security arena.

Bibliography

Alghamdi, G., Laskey, K. B., Wang, X., Barbara, D., Shackleford, T., Wright, E. J. and Fitzgerald, J. 2004. Detecting Threatening Behavior Using Bayesian Networks. Conference on Behavioral Representation in Modeling and Simulation - BRIMS, Arlington, VA, May 17-20, 2004.

CNN.com. 1998. Rationalizing Treason: An interview with Aldrich Ames. Cold War Experience - Espionage Series Retrieved January 20, 2005, from http://www.cnn.com/SPECIALS/cold.war/experience/spies/interviews/ames/.

CNN.com. 2001. The Case Against Robert Hanssen. from http://www.cnn.com/SPECIALS/2001/hanssen/.

Laskey, K. B. 2004, 2004/10/16. MEBN: Bayesian Logic for Open-World Reasoning. Retrieved Dec 8, 2004, from http://ite.gmu.edu/~klaskey/publications.html.